Clinical Trial Designs for the Early Clinical Development of Therapeutic Cancer Vaccines

By Richard M. Simon, Seth M. Steinberg, Michael Hamilton, Allan Hildesheim, Samir Khleif, Larry W. Kwak, Crystal L. Mackall, Jeffrey Schlom, Suzanne L. Topalian, and Jay A. Berzofsky

Abstract: **There are major differences between therapeutic tumor vaccines and chemotherapeutic agents that have important implications for the design of early clinical trials. Many vaccines are inherently safe and do not require phase I dose finding trials. Patients with advanced cancers and compromised immune systems are not good candidates for assessing either the toxicity or efficacy of therapeutic cancer vaccines. The rapid pace of development of new vaccine candidates and the variety of possible adjuvants and modifications in method of administration makes it important to use efficient designs for clinical screening and evaluation of vaccine regimens. We review the potential advantages of a wide range of clinical trial designs for the development of tumor vaccines. We address the role of immunological endpoints in early clinical trials of tumor vaccines, investigate the design implications of attempting to use disease stabilization as an end point and discuss the difficulties of reliably utilizing historical control data. Several conclusions for expediting the clinical development of effective cancer vaccines are proposed.**

J Clin Oncol 19:1848-1854.

DEVELOPMENT OF therapeutic cancer vaccines is a major area of oncologic research today. Whereas the important principles for the design of phase III trials apply to both tumor vaccines and chemotherapeutic drugs, there are major differences between these two classes of therapeutics that have important implications for early clinical development. First, the phase I concept of dose escalation to find a maximum-tolerated dose does not apply to most vaccines. Most vaccines are incapable of causing immediate serious or life-threatening toxicities at doses feasible to manufacture. Second, neither toxicity nor efficacy can be assessed in patients with advanced malignant disease associated with a blunted immune response because both toxicity and efficacy depend on the immune response. Consequently, issues of patient selection and end point definition require reconsideration for the early phases of vaccine development. Third, vaccination strategies often combine multiple agents, each of which needs to be optimized, such as adjuvants, cytokines, or costimulatory molecules. Although some of these combination regimens can be optimized in preclinical models, there remain comparisons that need to be performed in human patients. Thus, trial designs that allow the rapid screening of multiple variations on a regimen would greatly facilitate the development of cancer vaccines.

The purpose of this article is to address these and other issues and to provide approaches that facilitate the clinical development of therapeutic cancer vaccines. Early clinical development of tumor vaccines can be time consuming, costly, and ineffective. Much time can be wasted on unnecessary phase I trials or on phase II trials that ask phase III questions. We hope that this article helps investigators to design vaccine trials that are efficient, in terms of duration and numbers of patients treated, and yield valid and useful information to facilitate further vaccine development.

OBJECTIVES OF INITIAL VACCINE TRIALS

The initial use of an antitumor chemotherapeutic agent in humans is traditionally a dose-escalation phase I trial conducted in cancer patients with advanced metastatic disease that is refractory to other drugs. This trial design is rarely appropriate for tumor vaccines for the following reasons: (1) tumor vaccines are often based on DNA constructs, viral vectors, and cytokines that have been determined as safe from previous clinical trials; and (2) peptide vaccines generally seem inherently safe so long as the cytokine adjuvants are used in combinations and doses previously demonstrated to be safe.

For example, peptide vaccines based on nonmutated melanoma antigens such as MART-1/Melan A, and gp100 were initially evaluated in a phase I setting, at doses ranging from 0.1 to 10 mg. However, no toxicity was encountered even at the highest doses, and in vitro analysis did not reveal any correlation between peptide dose and the generation of specific T-cell reactivities from the peripherial blood lym-

From the Branches of Biometric Research, Environmental Epidemiology, Medicine, Pediatric Oncology, Surgery, and Metabolism; Biostatistics and Data Management Section; and Laboratory of Tumor Immunology and Biology, National Cancer Institute, Bethesda, MD.

Submitted July 27, 2000; accepted December 7, 2000.

Address reprint requests to Richard Simon, DSc, Chief, Biometric Research Branch, National Cancer Institute, Bldg EPN, Room 8134, Bethesda MD, 20892-7434; email: rsimon@nib.gov.

0732-183X/01/1906-1848

phocytes of vaccinated patients.[1,2] Thus, for subsequent trials using similar peptides, an intermediate fixed dose of 1 mg was chosen for vaccination, bypassing repetitive phase I analysis. Likewise, initial clinical trials using a novel virus or plasmid as a recombinant vaccine vector should be conducted in a phase I setting. However, if such vectors are proven to be nontoxic even at substantial doses, then subsequent trials using the same vectors but with different recombinant inserts may not require extensive dose escalation analysis. In addition, it should be noted that feasibility issues limit the maximum doses of certain recombinant proteins, viruses, or whole tumor cell vaccines that can be produced for administration to patients.

End-stage patients without intact immune systems may have very little likelihood of benefit or toxicity from a tumor vaccine. In some cases the potential toxicity of the regimen may be based on immune stimulation and will not be seen in anergic patients. Hence, such patients contribute little information about potential toxicities for individuals with intact immune systems.

The initial clinical trial of many new vaccines will not be a toxicity or dose-ranging trial but rather will involve administration of a fixed dose of vaccine to patients with relatively intact immune systems. The objective of such a trial is similar to those of most phase II trials, determination of whether the therapeutic agent shows sufficient biologic activity to warrant further development. In most cases, the dose selected will be based on preclinical findings or on practical considerations. Using several dose levels in the initial study to find the minimal active dose or to characterize the dose-activity relationship is generally not realistic. Although these may be desirable objectives, it may not be realistic to expect to accomplish them without studying a large number of patients. Finding the minimal active dose amounts to performing a conventional phase II trial at each of several dose levels. The fact that the trial is dose-ranging does not mean that three to six patients per dose level are adequate as in conventional toxicity trials. Those smaller sample sizes are only sufficient to exclude high toxicity rates. Phase II activity trials typically involve 14 to 25 patients in each dose level being evaluated. Characterization of the shape of the dose-activity relationship to select the smallest dose giving almost full activity is an even more ambitious objective. Such trials, if designed properly, require large sample sizes. Such trials may be more appropriate at later stages of development once the activity of the vaccine at some dose is established.

END POINTS

Clinical end points, such as tumor shrinkage, reduction in tumor marker, or delay in time to tumor progression, are more meaningful and interpretable than immunologic end points. However, traditional phase II trials in patients with clinically measurable tumors are often not appropriate as initial trials of tumor vaccines. Vaccine trials are best conducted in patients with intact immune systems, and for some diseases, this precludes inclusion of patients with gross tumor. As we discuss later, reliable assessment of whether a regimen delays recurrence of subclinical tumor cannot generally be accomplished by single-arm clinical trials. The designs we recommend for evaluating delay in tumor progression may be most effectively used after the basic immunogenicity of a vaccine regimen is established.

With rare exceptions,[3] tumor regression may not be obtainable with most vaccines in patients with advanced metastatic disease. The strategy of vaccination in a minimal disease state and using a sensitive tumor marker or molecular probe to measure reduction or disappearance of subclinical tumor mass can be an effective alternative for development of tumor vaccines. This approach was used in obtaining promising initial results for an idiotypic lymphoma vaccine which is now in randomized phase III testing using conventional clinical end points.[4] Although the relevance of a molecular marker of subclinical disease to long-term prognosis may be in question, such a marker can provide a measure of antitumor effect that can be measured in patients with minimal residual disease and is thus useful for early vaccine trials.

For some types of cancers, patients with clinically assessable tumors and intact immune systems are available, but existing vaccines do not produce clinical tumor responses. In such circumstances, immunologic end points may guide attempts to optimize the vaccine and its delivery. A phase III evaluation of a regimen in such a disease will rarely be warranted, however, until the regimen demonstrates activity based on a clinical end point.

For the above reasons, immunologic end points may be useful for early clinical trials of cancer vaccines. When an immunologic end point is used, we propose that the protocol should provide specific information about the variability of the measurement of that end point. Three sources of variability should be distinguished: variability among assay results on the same specimen (eg, lymphocytes, serum, or tumor tissue), variability among specimens from the same patient drawn at different times, and variability among patients. Documenting this variability is essential for interpretation of results of the trial. For example, such data permit one to define a threshold for change in the immunologic end point that can be regarded as statistically significant. If these data are not available at the outset, it may be possible to develop them during the course of the clinical trial (eg, from multiple baseline blood drawings for the

Table 1. Optimal Two-Stage Designs

Target Response Rate, p_1 (%)	No. of Patients		No. of Responses Required For Activity (A)	Probability of Early Termination (*P*)
	First Stage Sample Size, N_1	Maximum Sample Size, N		
20	12	37	4	.54
25	9	24	3	.63
30	7	21	3	.70
35	6	12	2	.74

patients to be vaccinated or from drawings for control patients). It is important to prepare plans for analysis of immunologic end points in advance and include these plans in the protocol to ensure that needed data are available and to reduce subjectivity in the analysis.

EFFICIENT PHASE II DESIGNS FOR SCREENING VACCINE REGIMENS

Knowledge of tumor immunology is advancing rapidly, and many vaccine regimens available for testing today will not be of interest in 2 years. There are many components of a vaccine regimen that may be altered such as route of administration, dosing schedule, and adjuvants used. There is also a wide range of approaches to tumor vaccination. Consequently, designs for efficiently screening tumor vaccines are needed. We will provide here some information on several designs that should be considered for phase II vaccine trials.

Optimal Two-Stage Designs

Simon's[5] optimal two-stage designs can be used to test whether a regimen has a response rate above a background level p_0. Frequently, $p_0 = 0.05$ is used. With clinical response, this assumes that no more than 5% of the patients will have apparent responses caused by variability in response assessment or spontaneous remissions. With immunologic end points, this design requires that immunologic response is defined at the outset and that an adequate estimate of end point reproducibility on repeat specimen sampling from the same patient is available. The two-stage design incorporates an early termination point, which allows the investigator to discontinue patient accrual if a desired end point has not been achieved in the first stage of the trial.

At the conclusion of the clinical trial, the regimen will be declared active or inactive. Table 1 lists several designs with 10% false-positive rate, 10% false-negative rate, and $p_0 = 0.05$. The false-positive rate (α) is the probability of declaring the regimen active when the true response probability is p_0. The false-negative rate (β) is the probability of declaring the regimen inactive when its true response probability is the target response rate p_1, the level of activity that we wish to be able to detect. In the first stage, N_1 assessable patients are entered and treated. If no responses are observed, then the trial is terminated and the regimen is declared inactive. Otherwise, accrual continues to a total of N assessable patients. At that point, accrual is complete. If the total number of responses is at least A, then the regimen is declared active. The last column of the table indicates the probability of early termination after the first stage when the true response probability is p_0. For example, if $p_0 = 5\%$ and the target response rate is 25%, then nine patients are treated in the first stage of the trial. If no responses are observed, the trial is terminated. Otherwise, accrual is continued to a total of 24 patients. If at least three responses are seen in the 24 patients, the regimen is declared active. The probability of declaring a regimen active when it's true response rate is 5% or less is 10%. The probability of missing the activity of a regimen with a true response rate of 25% is 10%. With a regimen having a response rate of 5%, the probability of stopping after only nine patients is 63%.

This design with $p_1 = 25\%$ and $p_0 = 5\%$ seems reasonable for many initial vaccine trials but designs based on other parameter values are easily generated using computer program OTSD (optimum two-stage design) available at http://lib.stat.cmu.edu/designs. The required number of patients depends strongly on the difference $p_1 - p_0$. These designs are based on a binary measure of response. We will deal with time to progression end points in a later section.

Randomized Phase II Trials to Select Among Experimental Regimens

Randomized phase II designs can be used to simultaneously screen several vaccine regimens. For example, suppose that we wish to separately evaluate three peptides whose activities are as yet unknown. One approach would be to conduct a three-arm randomized phase II trial. Randomization in this setting will eliminate the possibility of selection bias in patient accrual but should not be construed as providing phase III data. The accrual plan for each arm of the randomized design can be based on an optimal two-stage design as described in the previous section or can be determined to provide for selecting the most promising regimen among the arms.[6] Table 2 lists the roles in vaccine development for clinical trial designs we review here.

Several other designs have been published whose objectives are to select a promising treatment from among a large set of possible candidates. These methods include the approaches of Whitehead,[7] Strauss and Simon,[8] Yao et al[9] and Yao and Venkatraman.[10] The Strauss-Simon approach consists of a sequence of two-arm randomized phase II trials

Table 2. Features of Clinical Trial Designs for Vaccine Development

Design	End Point	Comparison	Role in Vaccine Development
Optimal two-stage	Dichotomous clinical, tumor marker, or immunologic response	None	Initial trial of fixed dose for establishing immunogenicity or antitumor activity
Randomized phase II	Dichotomous clinical, tumor marker, or immunologic response	None	Same as for optimal two-stage design but for several regimens simultaneously
Phase 2.5	Time to recurrence or response	Internal randomized control group	Initial trial for establishing antitumor activity in patients with subclinical tumor and intact immune systems
Factorial	Time to recurrence or response	Internally controlled for evaluating each factor	Screening several immunologic adjuvants or dosing schedules for optimizing an active vaccine

with the selected treatment from each trial carried over to the next trial, whereas the other approaches are nonrandomized. These approaches have helped document the trade-off between the sample size allocated for any particular clinical trial versus the number of treatments that can be studied in a fixed period of time. These types of designs have several limitations, however. (1) They provide only small sample sizes for each individual treatment when screening for large treatment effects. This can be accomplished more simply with the optimum two-stage design described in section 3 using large values of p_1-p_0. (2) With a very small sample size assigned to a particular treatment, one is very limited in the conclusions that can be drawn about that treatment. The results of the very small trial may not be publishable and a considerable amount of effort in filing Investigational New Drug Applications and obtaining IRB approvals will have been devoted to a rather inconclusive trial. Hence, this is a resource intensive approach. (3) The optimal sample sizes established from these designs may be sensitive to the assumed shape of the prior distribution of response probabilities of candidate treatments. As this distribution cannot be empirically estimated, there is considerable subjectivity in sample size determination.

HISTORICAL CONTROL COMPARISONS

In order to determine whether one vaccine regimen is more active than another, or whether it prolongs time to tumor progression compared to no vaccine treatment, a comparison is necessary. Several designs based on conducting a single arm trial and comparing results to those for patients treated previously on another trial have been described.[11-14] For example, this approach was used for a melanoma vaccine trial combining a peptide vaccine with high dose interleukin 2 (IL-2).[3] A large group of melanoma patients had been treated recently with high-dose IL-2 as a single agent by these investigators at the same institution, yielding $p_0 = 17\%$, which established a basis for evaluating the novel peptide vaccine combined with the identical IL-2 regimen in a phase II setting.

There are serious difficulties with all historical comparisons, however. Such comparisons are based on assumptions of comparability of patients, comparability of response assessments and comparability of ancillary care. Literature controls, remote historical controls or controls from other institutions are likely to be inadequate. Nonrandomized controls taken from previous trials are also likely to be biased when using immunologic end points because of changes in assay characteristics over time. Nonrandomized controls are very problematic for evaluating time to progression because intensity of tumor evaluation may vary. Contemporary controls based on a consecutive series of patients treated on a previous protocol at the same institution with a more objective measure of response are less likely to be problematic. Even so, comparability of prognostic factors must be carefully checked using individual patient data and bias in patient selection based on prior therapy or ability to tolerate the regimen must be scrutinized. In the best of circumstances, such contemporary controls will be reasonable for use in phase II type trials. In such cases, variability in estimate of activity of the historical control needs to be taken into account in planning the sample size of the single-arm trial.[13,14] From this perspective alone, unless there is a large series of patients that were contemporaneously treated, the use of a single-arm trial will not be effective. In any case, the ultimate test of the efficacy of any regimen must be validation in a randomized trial.

TIME TO PROGRESSION END POINTS

Evaluating whether a treatment delays recurrence or progression is particularly important for tumor vaccines. The importance of conducting initial vaccine trials in patients with intact immune systems was noted above and this may preclude use of patients with clinically measurable tumors. Evaluating the effect on a regimen on time to progression of subclinical disease is particularly problematic in a single-arm phase II trial, however. It is easy to devise a definition of disease stabilization, ie, lack of recurrence or progression for a specified period of time, but

the validity of the definition depends on the existence of data that establish that such stabilization does not occur in the absence of treatment. This is difficult to establish reliably because of the usual difficulties of identifying comparable nonrandomized controls and because of special difficulties involved with measuring time to disease progression in a consistent manner for different cohorts of patients. Consequently, we believe that the use of disease stabilization or time to progression as an end point in single arm trials should only be considered when data from a specific set of contemporaneous controls from the same institution are available. In such a case, rather than attempting to define disease stabilization in a valid manner as a dichotomous end point (eg, present or absent based on some threshold), it is preferable to compare the time to progression for the patients in the phase II trial to the distribution of time to progression of a specific set of control patients not receiving the vaccine regimen. Dixon and Simon[14] provide formulas for computing the number of patients required in the single arm trial.

Generally, we believe that single arm trials are not sufficiently reliable with use of time to progression endpoints. Instead, we prefer the randomized phase 2.5 design described below as a more satisfactory approach.

COMPARATIVE PHASE 2.5 TRIALS

Phase III trials are generally randomized comparisons of a new regimen compared to a standard treatment using an end point of established medical importance to the patient such as survival or quality of life. Phase III trials are usually planned using a 5% type one error parameter (α) because the results of phase III trials are viewed as definitive and are used as a basis for practice guidelines. We propose that in the development of cancer vaccines, there is a role for what might be called a phase 2.5 trial. Such a clinical trial would also be randomized, but may use an end point measuring biologic antitumor activity even though the end point might not be established as a valid surrogate for survival or quality of life. The phase 2.5 trial might also be based on an elevated statistical significance level since the objective of the trial would not be for regulatory approval or for establishing general practice guidelines.

To detect a large effect of a treatment in delaying tumor progression in a rapidly progressive disease such as pancreatic cancer or melanoma with visceral metastases does not require many patients in a randomized trial. With exponentially distributed times to progression, a 40% reduction in the hazard of progression corresponds to a 67% increase in median time to progression. To have 80% power ($\beta = 0.20$) for detecting this size of effect using a traditional $\alpha = 0.05$, only about 117 patients are required (assuming accrual rate of about 3 patients per month, median time to progression of 12 months for control group and follow-up time of 24 months after end of accrual)! This can be reduced to 87 patients if $\alpha = 0.10$. Hence, with 44 patients randomized to vaccine and the same number randomized to control, one can conduct a randomized phase 2.5 trial for evaluating whether the vaccine reduces the hazard of progression by 40%. This design would be a phase 2.5 design because of the unconventional use of a one-sided $\alpha = 0.10$ significance level and because time to progression might not be established as representing clear patient benefit. The phase 2.5 design is similar to the phase III design in the respect that it contains a control group for evaluating the experimental regimen and the intent is comparative. This is not the case for the randomized phase II design.

The basis of the efficiency is that the disease is rapidly progressive and a large treatment effect is targeted. If the median time to progression for the control group were 6 months instead of 12, an even smaller sample size would be required. When the disease is not rapidly progressive the efficiency illustrated here decreases. Statistical power for detecting a specified reduction of the hazard of an event is determined by the number of events, not the number of patients. With a slowly progressive disease, it may take many patients to be entered in order to observe a specified number of events unless the follow-up time following the close of accrual is very long. Also, for a rapidly progressive disease, a large reduction in hazard, eg, 40%, is associated with a moderate absolute increase in median time to event; eg, 6 months increased to 10 months, 12 months increased to 20 months. Consequently, with a rapidly progressive disease there is greater justification for targeting a relatively large treatment effect.

Although we believe that the randomized phase 2.5 design can play an important role in facilitating vaccine development, such clinical trials will generally not be adequate for regimen licensing. Phase 3 clinical trials with end points established as representing patient benefit and conventional 5% type one error rates will still be needed.

RANDOMIZED FACTORIAL TRIALS

Suppose one has an active vaccine regimen and now wants to evaluate two immunologic adjuvants (eg, IL-2 and granulocyte-macrophage colony-stimulating factor [GM-CSF]) to see whether either adds to the activity of the vaccine. One approach would be to conduct a two-arm randomized phase II design. The arms would be vaccine plus IL-2 and vaccine plus GM-CSF. This approach requires comparison to the historical control of vaccine alone, to evaluate whether either regimen looks promising relative to vaccine alone. Such a historical control comparison is

Table 3. 2 × 2 Factorial Design

	Response Rate (%)	
	Without IL-2	With IL-2
Without GM-CSF	15	35
With GM-CSF	15	35

problematic. A second approach would be to use a three-arm phase 2.5 design. The randomized control arm would be vaccine alone and the two other arms would be vaccine plus IL-2 and vaccine plus GM-CSF. A third approach would be to conduct a 2 × 2 factorial design.

With a 2 × 2 factorial design, there are four treatment groups. One group would receive vaccine alone; one would receive vaccine plus IL-2, one vaccine plus GM-CSF and one vaccine plus both IL-2 and GM-CSF. Patients are randomized to the four groups. For analyzing the effect of IL-2, the response rate for the two arms which received IL-2 are compared to the response rate for the two arms which did not receive IL-2. Analyzing the effect of GM-CSF is similar. The factorial design can be considered either a phase 2.5 or a phase 3 design depending on the type one error level (α) used and whether the end point used is of demonstrated patient benefit.

The factorial design has some efficiency advantages over the three-arm phase 2.5 design described earlier. In a 2 × 2 factorial design, each factor is evaluated by comparing outcomes for all patients receiving that factor to outcomes for all patients not receiving that factor. Therefore, the sample size of a factorial design is usually the same as for a simple two-arm trial.[15] Thus, a 2 × 2 factorial design for studying the vaccine with or without GM-CSF and with or without IL-2 will require only two thirds the number of patients as a three arm trial comparing vaccine alone to vaccine plus GM-CSF and vaccine plus IL-2. This assumes that the same design parameters are used in both cases.

Consider, for example, the parameters listed in Table 3. The response rate of the vaccine alone is 15%, GM-CSF is not effective, but IL-2 increases the response rate to 35%. In order to detect this with 90% statistical power and $\alpha = 0.10$ (one-tailed), 70 patients per group are required for a two group comparison. Since each of these two groups is a composite of patients receiving and not receiving GM-CSF, the factorial trial would require 35 patients per arm for each of the four arms. If we wish to allow for the possibility that both IL-2 and GM-CSF are effective adjuvants, the sample size per arm would have to be increased somewhat. The sample size is determined based on evaluating the treatment effect of a factor computed within each level of the other factor and averaged over those levels. This will provide the ability to obtain comparative conclusions about the contributions of GM-CSF and IL-2. The sample size is not adequate for comparing the arms individually with each other; eg, for comparing vaccine plus IL-2 versus vaccine plus GM-CSF.

Early stopping rules can be used with a factorial design. For example, if no responses are seen in the first 12 patients on the IL-2 arms, one can conclude that the IL-2 containing arms do not on average provide for a 25% response rate. Similarly, this is the case for the GM-CSF arms. Consequently, if no responses are seen after 24 patients are entered on trial (six patients per arm), the entire trial could be stopped. Otherwise, the trial would go to a maximum of 140 patients.

Factorial designs scale very well. The total sample size required does not increase as the number of factors increase. Factorial designs are also very efficient and internally controlled for screening several immunologic adjuvants for an effect on time to progression using patients with intact immune systems.

The main concern expressed about factorial designs is that there could be interactions among the factors. That is, the effect of IL-2 may differ depending on whether or not GM-CSF is administered. If biologic interactions that change the effects of the factors on response are expected, then the factorial design may not be appropriate. While biologic interactions are likely with such molecules, it often works out that neither factor has any effect on response. Hence, the factorial design is often an efficient approach for screening ineffective factors. The IL-2 effect tested by a factorial design is the average of the effect without GM-CSF and the effect with GM-CSF. For Table 3 we have assumed that this average effect corresponds to a 20-percentage point difference in response probabilities. So long as the average effect is of this size, the planned statistical power will be at least approximately achieved. Synergistic interactions that maintain or enhance the targeted average effects do not reduce the power of factorial designs to detect main effects. In fact, factorial designs are effective mechanisms of screening for combinations of factors that are synergistic for response although the power for definitively testing for the presence of interactions will generally be low.

DISCUSSION

Therapeutic cancer vaccine development is a broad field and different clinical trial designs will be appropriate in different circumstances. Nevertheless, we believe that there are several general recommendations that can be made.

Traditional phase I designs based on escalation from a very low starting dose in patients with advanced cancer are not always necessary in cancer vaccine development. It may

be more relevant to treat patients with intact immune systems and less advanced cancers, with either a single dose or a very limited number of escalation steps.

Clinical trials are necessarily less efficient than laboratory experiments and as much optimization as possible of vaccine schedule, mode of delivery, adjuvants and cofactor molecules should be conducted in preclinical models.

Immunologic endpoints are problematic because of assay variability and uncertainty of clinical relevance. Immunologic end points should generally be utilized only to establish a basic biologic effect of the regimen unless the mechanism of immunologic action is well understood. When an immunologic end point is used, assay reproducibility should be carefully documented and immunologic response criteria carefully established based on statistical considerations.

Identification of activity can be accomplished efficiently using a traditional two-stage phase II design with $p_0=0.05$, $p_1=0.25$ and α and β errors of 0.10. This design can be used as the initial design using an immunologic response end point.

When there are several candidate regimens available for screening, use of a randomized phase II design helps avoid selection bias in assigning new patients to the regimens. Each arm of the trial can be sized as if it were a single arm two-stage phase II design with $p_0=0.05$, $p_1=0.25$ and α and β errors of 0.10 because the objective of such a trial is not comparative.

Comparison of a new vaccine regimen to a control group from a previous trial is hazardous. In exceptional circumstances with a large series of control patients available from a contemporaneous protocol at the same institution with the same eligibility criteria and response assessment as in the new protocol, this approach may be useful. Even then, careful evaluation of comparability should be performed and results of the comparison used for limited phase II purposes, not for claiming treatment effectiveness.

Randomized phase 2.5 trials of vaccines using time to disease progression or recurrence as end point in rapidly progressive diseases provide reliable and efficient designs for screening vaccines.

Randomized factorial designs provide an efficient tool for screening multiple factor modifications. This design is well suited to factors, which are not toxic. As long as the factors do not interact in therapeutically significant ways, this design can be used for simultaneous screening of multiple factors.

ACKNOWLEDGMENT

We thank the members of the National Cancer Institute Vaccine Working Group, Dr Edison Liu, and Dr Richard Klausner for active discussions and helpful suggestions.

REFERENCES

1. Salgaller ML, Marincola F, Cormier JN, et al: Immunization against epitopes in the human melanoma antigen gp100 following patient immunization with synthetic peptides. Cancer Res 56:4749-4757, 1996

2. Cormier JN, Salgaller ML, Prevette T, et al: Enhancement of cellular immunity in melanoma patients immunized with a peptide form MART-1/Melan A. Cancer J Sci Am 3:37-44, 1997

3. Rosenberg SA, Yang JC, Schwartzentruber DJ, et al: Immunologic and therapeutic evaluation of a synthetic peptide vaccine for the treatment of patients with metastatic melanoma. Nat Med 4:321-327, 1998

4. Bendandi M, Gocke CD, Kobrin CB, et al: Complete molecular remissions induced by patient specific vaccination plus granulocyte-monocyte colony stimulating factor against lymphoma. Nat Med 5:1171-1177, 1999

5. Simon R: Optimal two-stage designs for phase II clinical trials. Control Clin Trials 10:1-10, 1989

6. Simon R, Wittes RE, Ellenberg SS: Randomized phase II clinical trials. Cancer Treatment Rep 69:1375-1381, 1985

7. Whitehead J: Designing phase II studies in the context of a program of clinical research. Biometrics 41:373-383, 1985

8. Strauss N, Simon R: Investigating a sequence of randomized phase II trials to discover promising treatments. Stat Med 14:1479-1489, 1995

9. Yao TJ, Begg CB, Livingston PO: Optimal sample size for a series of pilot trials of new agents. Biometrics 52:992-1001, 1996

10. Yao TJ, Venkatraman ES: Optimal two-stage design for a series of pilot trials of new agents. Biometrics 54:1183-1189, 1998

11. Thall PF, Simon R: A Bayesian approach to establishing sample size and monitoring criteria for phase II clinical trials. Control Clin Trials 15:463-481, 1994

12. Thall PF, Simon R, Estey E: Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes. Stat Med 14:357-379, 1995

13. Makuch RW, Simon R: Sample size considerations for nonrandomized comparative studies. J Chronic Dis 33:171-175, 1980

14. Dixon DO, Simon R: Sample size considerations for studies comparing survival curves using historical controls. J Clin Epidemiol 41:1209-1213, 1988

15. Simon R, Freedman LS: Bayesian design and analysis of 2 by 2 factorial clinical trials. Biometrics 53:456-464, 1997